

# 빅데이터 분석을 이용한 지하철 혼잡도 예측 및 추천시스템

김진수  
안양대학교 교양대학

## Subway Congestion Prediction and Recommendation System using Big Data Analysis

Jin-su Kim  
College of Liberal Arts, Anyang University

**요 약** 지하철은 버스와 택시에 비해 많은 승객들을 안전하고 신속하게 대량 수송할 수 있는 미래 지향적인 교통수단이다. 지하철 이용자의 증가에 따른 혼잡도 증가는 지하철을 쾌적하게 이용할 수 있는 시민들의 권리를 저해하는 요인 중의 하나이다. 따라서 지하철 내의 혼잡도 예측은 승객의 이용 편의성과 쾌적성을 극대화할 수 방법 중 하나이다. 본 논문에서는 기존의 지하철 혼잡도를 다중 회귀 분석으로 예측하고 빅데이터 처리를 통한 실시간으로 혼잡도를 모니터링하고, 자신의 출발역과 도착역 정보뿐만 아니라 다양한 정보를 추가하여 개인화된 혼잡도 예측 시스템을 제안한다. 제안된 혼잡도 예측 시스템을 적용한 결과 예측혼잡도가 실제혼잡도에 비해 평균 81% 정확도를 보였다. 본 논문에서 제안한 예측 및 추천 어플리케이션을 지하철 고객에 적용하면 지하철 혼잡도 예측과 개인 사용자의 편리성에 도움이 될 것으로 예상된다.

**주제어** : 혼잡도 예측, 빅데이터, 다중 회귀 분석, 혼잡도 예측 앱, 개인 선호도

**Abstract** Subway is a future-oriented means of transportation that can be safely and quickly mass transport many passengers than buses and taxis. Congestion growth due to the increase of the metro users is one of the factors that hinder citizens' rights to comfortably use the subway. Accordingly, congestion prediction in the subway is one of the ways to maximize the use of passenger convenience and comfort. In this paper, we monitor the level of congestion in real time via the existing congestion on the metro using multiple regression analysis and big data processing, as well as their departure station and arrival station information. More information about the transfer stations offer a personalized congestion prediction system. The accuracy of the predicted congestion shows about 81% accuracy, which is compared to the real congestion. In this paper, the proposed prediction and recommendation application will be a help to prediction of subway congestion and user convenience.

**Key Words** : Congestion Prediction, Big Data, Multiple Regression Analysis, Congestion Prediction Application, Personal Preference

Received 2 October 2016, Revised 1 November 2016  
Accepted 20 November 2016, Published 28 November 2016  
Corresponding Author: Jin-Su Kim  
(College of Liberal Arts, Anyang University)  
Email: kjspace@anyang.ac.kr

ISSN: 1738-1916

© The Society of Digital Policy & Management. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

www.kci.go.kr

## 1. 서론

지하철은 버스와 택시와 같은 교통수단에 비해 많은 승객들을 원거리까지 안전하고 신속하며 정확하게 원하는 지점으로 대량 수송할 수 있어 미래 지향적인 교통수단이라 할 수 있다. 이러한 지하철 운행 특성에 따라 출퇴근 시간뿐만 아니라 일상적인 대중교통으로써의 수요 또한 급증하고 있다. 1995년 134만 명을 기준으로 2010년 236만 명으로 약 100만 명이 꾸준히 증가하고 있고, 6대 도시권에서 1시간 이상 출퇴근하는 이동 인구가 76% 증가하였다[1]. 지하철 이용자의 증가에 따라 발생하는 혼잡도를 줄이기 위한 연구는 쾌적한 지하철(철도)을 이용할 수 있도록 시민들의 권리를 찾는 교통복지와 안전 차원의 기초 연구 중 하나이다[2,3]. 교통복지 개념은 운행비용, 통행시간, 사고비용, 환경비용, 차내 혼잡도, 주차비용, 접근성, 정시성, 교통약자시설의 이용 편의 등으로 구성되며, 이 중 가장 높은 비중의 지표가 혼잡도이다[4,5].

승객 혼잡도를 해결하고 승객의 안전과 편의성을 제공하기 위해 CCTV를 이용하여 혼잡도를 측정하고, 공공데이터 분석을 통해 새로운 대안을 제시하는 등의 노력도 있지만[6], 이보다 앞서 개인정보나 초상권등과 같은 인권이 보호되는 범위 내에서 보장되도록 해야 한다.

본 논문에서는 지하철 내의 혼잡도 예측을 통한 승객의 이용 편의성을 극대화하고, 군중밀도가 높은 지하철 구간의 경우 차량 칸별 고른 승객 분포를 유도하여 승객의 편의성을 높일 수 있는 개인 혼잡도 예측 시스템을 제안한다. 본 논문에서 제안하는 시스템은 기존의 지하철 혼잡도와 빅데이터 처리를 통한 실시간 혼잡 상황을 모니터링한다. 모니터링된 정보와 개인의 출발역과 도착역 정보를 기반으로 출발역과 도착역(환승역 포함)의 다양한 정보를 예측 및 추천한다. 개인화된 정보와 개인의 선호도를 적용하여 전동차 각 구간의 혼잡도 정보를 미리 예측한다. 예측된 혼잡도를 바탕으로 효율적인 추천을 유도할 뿐만 아니라 전체적인 혼잡도를 균등하게 분포시킬 수 있다. 또한 승강장 및 지하철 객차내의 혼잡도는 승객들의 동선을 최소화하려는 심리적 요인으로 인해 환승통로, 환승계단, 환승 엘리베이터 등과 같은 특정지점이 타 지점에 비해 혼잡도가 상대적으로 높다. 이러한 심리적 요인과 승객의 출발역, 도착역, 환승역 정보를 파악

하여 보다 균등한 객차 혼잡비율을 조절할 수 있는 객차 번호를 추천하고자 한다.

## 2. 관련연구

### 2.1 혼잡도 지표

버스는 단독으로 통행하거나 동일 교통수단인 버스끼리 환승할 경우에 승차 및 하차에 대한 정보가 동일하게 관리되어 혼잡도 예측을 위한 지점 및 노선별 환승의 규모나 유형을 분석하기 용이하나, 지하철(철도)은 다양한 교통수단의 정보가 혼재하여 혼잡도를 예측하기가 버스의 경우와는 다르다. 민자 노선과 같이 환승게이트가 설치된 경우를 제외하고는 대부분의 지하철 노선에서 환승게이트 없이 환승하기 때문에 지하철간의 환승은 제대로 파악되지 않아 해당 역의 승하차 인원만으로는 혼잡도를 정확히 예측하기가 어렵다.

혼잡도 지표는 지하철 운행 시 승객들의 안전 및 복지를 위한 가장 중요한 요인 중 하나이다. 국토교통부내 도시철도 건설과 지원에 관한 기준에 나타난 자료에는 혼잡도가 150%를 넘는 경우 해당 노선에 증량 또는 증편을 요구하고 있기 때문에, 운영기관에서는 출퇴근시간 및 혼잡 시간대를 분석하여 열차 간 시간 조정 자료 활용과 증량, 증편 등의 정책 결정 자료로 이용된다. 수도권역의 여러 운영기관에서는 혼잡도를 1년이나 2년 주기로 발표하고 있으며, 차량 증편이나 수요대응 열차 스케줄에 활용하고 있다. 그러나 현재 사용 중인 혼잡도 측정은 대부분 목적측 방법과 설정된 경로 데이터를 활용하여 추정하는 것이다. 이러한 방법은 단일구간이나 환승역이 상대적으로 적은 경우에 적용하기 용이하며 복잡한 수도권 망에서의 적용은 오차가 매우 클 뿐만 아니라 운영기관이 자체 측정 및 발표로 인한 중립성 논란과 객관적 혼잡도 추정을 위한 신뢰성을 높일 수 있는 다양한 방법이 필요하다.

철도운영기관이 혼잡도를 중시하는 이유는 효율적인 차량 운영 계획 때문이다. 각 운영기관에서 혼잡도의 활용은 첫째, 환승현황 및 혼잡도 결과를 토대로 제반도시철도 시설물의 안전관련 사항과 역운영의 참고자료, 둘째, 도시철도를 이용하는 승객에 대한 안전 확보와 서비스 향상 자료, 그리고 세번째는 교통량의 변동량 추이를

분석하여 승객 서비스 수준을 유지하기 위한 자료 등이다. 대부분의 운영기관에서 사용하는 혼잡도 추정 방법은 승하차실적(인·km)과 목측 방법을 주로 사용한다. 기관별로 운영되는 차량 규모에 따라 다소 계산방식이 다르지만 방법론적으로 보면 큰 차이는 없다. 인천교통공사에서 이용하는 혼잡도(Congestion) 산정 기준은 다음과 같다.

$$\text{Congestion} = \text{Passengers} / \text{Capacity} \times 100 \quad (1)$$

여기서, Passengers는 총 재차인원으로 좌석인원과 입석인원의 총합이고, Capacity는 차량 정원을 의미한다.

그러나 재차인원을 정확히 계수하기가 어렵기 때문에 목측기준을 이용하여 측정하는데 혼잡도 추정 조건은 다음 <Table 1>과 같다[7].

<Table 1> Eye Measurement of Congestion in Incheon Transit Corporation

#Passenger (person)	Congestion (%)	Passenger distribution in Subway
24	19.4	50% sit
48	38.7	100% sit
80	64.5	100% sit, 50% handle
96	77.4	100% sit, 50% handle, 2 person on each door
112	90.3	100% sit, 100% handle
124	100.0	100% sit, 100% handle, 2 person on each door
158	127.4	100% sit, 100% handle, 2 person on each door, 1 line in middle
238	191.9	100% sit, 100% handle, 10 person on each door, 1 line in middle
272	219.4	100% sit, 100% handle, 10 person on each door, 2 line in middle, the passengers are in close contact with each door
300	241.9	subway saturation. Failure some passengers boarding

## 2.2 빅데이터(Big Data)

인터넷 사용의 대중화, 정보기술의 비약적인 발전, 그리고 다양한 모바일 기기의 급속한 보급을 통해 사용자가 원하는 정보를 좀 더 용이하고 빠르게 처리할 수 있는 환경을 제공한다. 다양한 형태의 정보기술 활용 중 각종 소셜 네트워크 서비스(Social Network Service, SNS)를 통해 자신의 정보를 단순히 공유하고 교환할 뿐만 아니라 SNS를 통해 수집된 다양한 정보를 개인 혹은 집단 지성 축적을 위해 정보의 재생산, 재분석, 재배포 등을 수행하고 있다. 이와 같은 정보기술의 발전과 이용 확산은 데

이터의 양적 팽창을 가져왔고, 여러 분야에서 데이터의 흐름을 주시할 필요성이 제기되었다[8].

빅데이터(Big Data)는 기존 데이터베이스 관리 도구의 역량을 넘어서는 대량의 정형 또는 비정형의 데이터 집합으로부터 의미있는 가치를 추출하고 결과를 분석하는 기술이다[9,10]. 빅데이터의 특징을 양(Volume), 다양성(Variety), 속도(Velocity)의 3V로 표현한다. 다양한 공간에 산재된 방대한 데이터는 유용한 정보뿐만 아니라 불필요한 정보도 동시에 기하급수적으로 증가하기 때문에 의미를 지닌 지식을 찾아내기 위한 빅데이터 분석기법들이 필요하다. 이러한 빅데이터 분석기법을 이용하여 소셜 네트워크 분석, 유전체 분석, 고객 데이터 관계 분석, 실시간 사물지능통신 등 다양한 분야에 활용된다. 빅데이터 분석기술에는 텍스트 마이닝, 오피니언 마이닝, 클러스터 분석, 데이터 마이닝, 소셜 네트워크 분석 등이 있으며[11], 빅데이터 분석을 이용하여 공공부문, 비즈니스 마케팅, 개인의 소셜 네트워크 분석 등 다양한 분야에서 다양한 방법론으로 연구가 진행되고 있다[12,13].

## 3. 혼잡도 예측 및 추천 시스템

본 논문에서 설계한 개인 선호도에 따른 혼잡도 예측 및 추천 시스템은 먼저 각 지하철역에 대한 다양한 정보 데이터베이스를 생성하고, 각 역별 혼잡도 예측을 위한 다중 회귀 모형을 구축한다. 사용자가 출발역과 도착역, 최적/최단 선택을 설정하면 개인화된 정보에 따른 예측된 혼잡도와 승하차(환승 포함) 지점에 대한 다양한 정보를 추천한다.

### 3.1 지하철역 관련 DB 구축

본 논문에서는 지하철 혼잡도 예측을 위해 2013년과 2014년도 통계자료를 수집하여 데이터베이스에 저장한다. 또한 개인화된 승객의 개인적인 요구 사항을 반영하기 위해 각 플랫폼의 계단, 장애인 및 노약자용 엘리베이터 등의 위치 정보, 차량의 칸번호, 문번호와의 매칭을 위한 데이터베이스를 구축한다. 예를 들어, 환승역이라면 환승통로와 가장 빠른 차량의 칸 번호와 문 번호를 저장한다. 비환승역이라면 계단, 엘리베이터 등과의 최단거리에 있는 칸번호와 문번호를 저장하며, 각 역 내의 편의시

〈Table 2〉 Independent variables and data

Station	# In Passenger	# Out Passenger	# Exit	# All Access	# Transit Stop	Income	# Building	# BigData	Waiting Time	Bus/Subway	Congestion
Gyeyang	1,399,358	1,388,394	1	2	16	1,187,131	9	495,550	25	10	16.9
Bakchon	1,612,481	1,463,739	4	5	12	1,409,893	28	525,680	11	10	14.6
Jakjeon	5,727,692	5,549,084	8	9	36	5,251,807	65	961,640	13	47	62.8
Bupyeong Market	4,195,566	4,245,443	4	5	57	3,477,720	30	572,960	11	39	86.5
Bupyeong	2,565,969	2,775,460	8	10	57	2,262,085	66	566,000	13	67	63.6
Ganseogogeor	4,889,324	4,719,153	9	10	41	3,984,783	35	340,800	21	36	76.1
Arts Center	4,205,742	4,313,187	11	12	71	3,893,360	64	528,200	28	18	78.9
Sinyeonsu	2,285,105	2,194,058	4	6	23	2,048,949	29	461,150	20	10	65.4
Woninjae	1,412,913	1,313,947	6	7	28	1,347,747	57	397,860	15	7	58.2
Dongmak	2,273,319	2,195,153	4	6	31	2,024,592	21	493,330	19	13	25.6
Campus Town	1,624,492	1,569,299	3	11	18	1,563,976	17	322,800	20	14	23.1
Central Park	749,579	762,228	3	7	14	772,865	10	498,080	9	3	1.0

설, 출발역과 도착역(환승역)사이의 소요시간 등의 시간 정보도 데이터베이스화한다.

### 3.2 빅데이터 수집 및 개인 선호도 조사

SNS나 인터넷으로부터 지하철역 및 혼잡도와 연관된 단어들의 출현빈도를 추출한다. SNS에 지하철역명이 출현하면 SNS 사용자가 해당 역에 위치한다고 가정하고 해당 역과 함께 사용된 단어들을 분석하여 해당 역 근처에 대중이 모일 수 있는 원인을 추출, 유추하고 해당 사항에 대해 가중치를 적용하여 혼잡도 예측에 사용한다. SNS에 나타난 해쉬 태그는 역과 연관된 중요한 키워드이므로 역명과 해쉬 태그와의 연관도를 분석하여 개인의 선호도를 조사한다.

### 3.3 다중 회귀 모형을 이용한 독립 변수 및 파생 변수 추출

다중 회귀 분석[14]은 두 개 이상의 독립 변수들에 대한 종속 변수인 다른 변수들 사이의 결합분포인 함수 관계를 수학적 모형을 통해 표현하여 인과관계를 분석한다. 또한 도출된 수식을 통해 독립 변수들의 변화에 따른 종속변수의 변화를 예측 가능하도록 하는 통계기법이다. 다음은 독립변수가 k개인 선형 회귀 모형을 표현한 것이다.

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki} + \epsilon_i, i = 1, 2, \dots, n \quad (2)$$

여기서  $\beta_0, \beta_1, \dots, \beta_k$ 는 모집단의 회귀계수이고,  $\epsilon_i$ 는 종속 변수  $y$ 를 계산할 때 발생하는 오차이다. 이때, 오차  $\epsilon_i$ 는

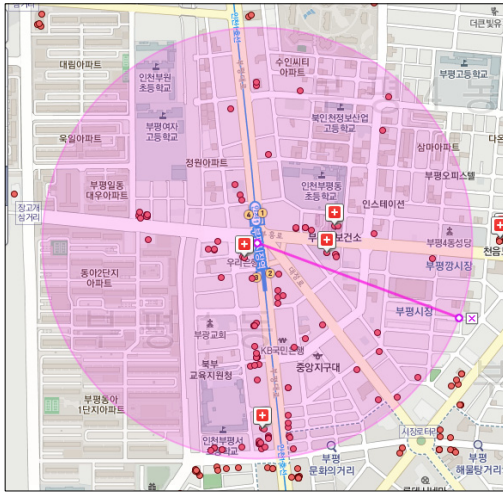
정규분포  $N(0, \sigma^2)$ 을 따르며 오차들 서로 독립이라 가정한다.

혼잡도 예측을 위한 독립변수에는 통계데이터 및 관측 자료로 수집한 승차여객수(# In Passenger), 하차여객수(# Out Passenger), 여객수입(Income)과, SNS에 검색된 게시물 수(# BigData), 500m 이내의 환승 정류장 수(# Transit Stop), 500m 이내의 버스 노선수(# Bus), 열차당 버스 수(Bus/Subway), 열차 당 버스 예상 간격 시간(wait), 500m이내의 편의 시설(의료시설, 백화점, 쇼핑몰 등) 수(# Building) 등이다. 이러한 다중회귀분석에 사용되는 독립변수 후보들과 후보들의 관측된 값들을 <Table 2>에서 보여준다.

## 4. 실험 및 분석

개인 선호도에 따른 혼잡도 예측 및 추천을 위해 인천 지하철1호선의 역을 모집단으로 선정하여 자료를 구축하고 실험하였다. 인천지하철 1호선의 경우 전동차의 규모가 작은 중량 열차가 운행되고 있으며, 하나의 편성당 차량 대수도 8량이다. 차량 정원은 좌석과 입석을 모두 포함하여 총 970명으로 승객 수송 능력은 타 노선에 비해 상대적으로 작다. 국가통계포털에서 인천광역시 기본 통계 중 지하철 수송(2012~2014) 통계조치를 통해 자료를 수집하였고, 2013년 인천의 1일 시민교통량 중 12.2%인 약 738,000명이 지하철을 이용한다[15].

<Table 2>는 종속변수인 혼잡도에 영향을 주는 독립 변수들과 자료들을 보여주며, <Table 2>에 나타난 혼잡



[Fig. 1] The buildings around the station within 500 meters

도는 국제업무지구방향의 첨두시간에 측정된 것을 사용하였다[16].

독립 변수들에 해당하는 자료 수집은 통계청 자료와 웹문서 및 SNS 내의 혼잡 관련 내용 탐색, 그리고 다음 지도로부터 일정 반경 내에 있는 버스 정류장 정보, 대형 건물, 버스 시간 정보 등의 다양한 자료를 수집하였다.

다중 회귀분석을 통해 혼잡도에 영향을 주는 주요한 변수들을 추출한 결과, 수입(천원)을 제외한 승하차여객수(# In Passenger, # Out Passenger), 출입구(# Exit), 통로(# All Access), 환승 정류장수(# Transit Stop), 건물수(# Building), 빅데이터(# Big Data), 열차당 버스수(Bus/Subway) 등이 95%의 신뢰도를 기준으로  $p < 0.05$ 인 유효한 결과로 계산되었다. <Table 3>은 다중회귀분석의 결과를 보여준다. 혼잡도 타당성 검증을 위해 도출된 회귀식에 인천지하철1호선의 각 역에서 관측된 독립

<Table 3> Result of Multiple Regression Analysis

	Coefficient	Std. error	T statistic	P-value
Y	127.0099	16.89461	7.517776	0.004878
# In Passenger	-0.00026	5.73E-05	-4.48308	0.020699
# Out Passenger	0.000277	6.1E-05	4.551246	0.019875
# Exit	12.76019	1.981031	6.441185	0.007588
# Full Access	-13.7801	1.671303	-8.24513	0.003735
# Transit stop	-2.15504	0.549496	-3.92184	0.029489
# Big Building	14.00625	1.622741	8.631229	0.003271
# Big Data	-0.0002	2.34E-05	-8.57821	0.00333
Bus/subway	-1.04727	0.196522	-5.32904	0.012913

변수들의 값을 적용했을 때 예측된 혼잡도와 실제 관측된 혼잡도, 그리고 예측 정확도는 <Table 4>에서 보여주며, 예측혼잡도는 실제혼잡도와 비교하여 평균 81%의 정확도를 보였다. 그러나 테스트한 역 중의 혼잡도 예측과 실제 혼잡도 차이가 현저히 나는 경우도 발생하였다. 이는 부평역과 같은 많은 승객의 환승이 발생하는 허브역을 중심으로 허브역 방향에 가까운 역일수록 각 역에서 추출한 승하차 인원 상관계수가 혼잡도가 지속적으로 증가하였다.

혼잡도를 계산하기 위한 식(3), 식(4)에서  $i$ 는 현재 역,  $i-1$ 은 이전 역을 의미하고,  $CI_{i-1}$ 는 이전역의 혼잡도,  $\alpha_i, \beta_i$ 는 각각 현재역의 승하차 인원,  $T\alpha_i, T\beta_i$ 는 환승역의 경우 환승하는 승하차 인원, 그리고  $N$ 은 차량정원이다.

$$CI_i = (CI_{i-1} \times N + \alpha_i - \beta_i + T\alpha_i - T\beta_i) \times \gamma(i) / N \times 100 \quad (3)$$

$$\gamma(i) = \frac{2}{1 + e^{-x}} \quad (4)$$

단, 사고나 행사와 같은 특별한 시점에 나타날 수 있는 사건을 SNS와 같은 매체를 통해 출현한 빈도수( $x$ )를 이용하여 가중치  $\gamma(i)$ 를 결정한다. 가중치는 [1, 2] 내의 값으로 설정된다. 즉, 가중치 1은 빅데이터에 출현한 빈도

<Table 4> Comparisons between Predict Congestion and Real Congestion

Station	# In Passenger	# Out Passenger	# Exit	# All Access	# Transit stop	# Building	# Big Data	Bus / Subway	Predict Congestion	Reall Congestion	Accuracy
Gyesan	4,783,054	4,665,250	6	7	32	10	743,500	31	62	50	80%
Bupyeong-gu Office	2,611,519	2,739,115	9	14	52	16	491,500	50	99	82	83%
Incheon City Hall	2,536,563	2,515,677	9	10	41	9	541,300	19	60	48	81%
Seonhak	2,676,364	2,625,442	4	5	18	5	506,770	12	67	56	84%
Bupyeong samgeori	986,883	952,894	4	5	36	3	252,200	12	21	16	75%

수  $k$ 가 0일 때이며, 무한대에 가까운 출현빈도일 경우 가중치는 2이다.

혼잡도는 침두시간과 비침두시간에 따라 다르게 측정되며, 출퇴근 침두시간에 따라 변하는 특징이 있다. 또한, 인천지하철1호선의 경우 부평역 방향(허브역)으로 갈수록 혼잡도가 증가하며 부평역 반대방향으로 갈수록 혼잡도가 감소하는 경향이 있다. 본 논문에서 제안한 혼잡도 예측 시스템을 통해 개인화된 대체와 연동하여 개인화된 선호도에 따라 혼잡한 차량 내 사용자별 효율적인 차량 및 칸을 선택할 수 있다. 먼저, 각 사용자의 출발역과 도착역을 설정, 개인이 원하는 최단/최적 선택을 지정하면 사용자가 입력한 시간과 도착 시간을 기준으로 예측되는 혼잡도와 도착지 또는 환승역 차량 칸의 위치 및 문 번호를 추천해 보여준다. [Fig. 2]는 스마트폰에 구현된 최적/최단 혼잡도 추천 시스템의 다양한 정보를 보여준 화면이며, 본 논문에서 제안한 개인 선호도에 따른 혼잡도 추천 시스템 앱을 사용한 실험자 30명에 대한 테스트한 결과 23명이 만족한 결과를 보였다.



[Fig. 2] Example of Congestion Prediction and Recommendation

## 5. 결론 및 향후 연구

본 논문에서 제안한 혼잡도 예측 및 추천시스템은 각 지하철역에서 수집한 통계자료와 SNS 실시간 자료를 포함한 주변 정보를 활용하여 다중회귀모형을 구축한다. 사용자가 이동하려는 역정보를 앱에 설정하면, 사용자의 위치, 시간 등의 정보와 현재의 빅데이터 분석 결과에 따른 예상 혼잡도 및 다양한 추천 정보를 보여준다. 본 시스템의 예측 혼잡도는 약 81%정도의 정확도를 보였다.

그러나 특정 환승역과 같은 대규모 중심역 방향쪽의 역들은 낮은 예측도에 비해 혼잡도가 증가되고, 반대 방향은 혼잡도가 감소되었다. 이러한 혼잡도 예측의 오류를 보정하기 위해서 허브역 방향에 가까울수록 혼잡도 가중치를 적용해야 할 것이다.

향후 연구에서는 이러한 허브역을 기준으로 가중치를 부여하여 혼잡도 예측에 적용하면 보다 효율적인 예측이 가능할 것이며 역 주위의 사건 사고나 행사 등과 같은 다양한 혼잡도 예측 인자 등에 대해서도 연구해야 할 것이다.

## REFERENCES

- [1] Yong-Hyun Cho, "Metropolitan commuting time in half", Koera Railroad Research Institute, 2013.
- [2] Seung-Kirl Baek, Dong-Joo Park, "A Status and Policy Direction of Transportation Welfare in Inter-regional Transportation", Transportation technology and policy, v.9, no.5, pp.52-60. 2012.
- [3] Seon-Ha Lee, Choon-Keun Cheon, Byung-Doo Jung, Byung-Young Yu, Eun-Ji Kim, "Study on Methodology for Effect Evaluation of Information Offering to Rail passengers - Focusing on the Gate Metering Case Study considering congested conditions at a platform -", The Korea Institute of Intelligent Transport Systems, Vol.14, No.3, pp.50-62, 2015
- [4] Jong-Hyung Kim, Jhi-Eon Sohn, The Congestion Index of Urban Rail for the Transportation Welfare in Incheon, Incheon Development Institute, 2014.
- [5] Mi-Young Bin et al., Study on Rail Construction Feasibility by Considering Transportation Welfare, Gyeonggi Research Institute, 2012.
- [6] Keun-Won Kim, Dong-Woo Kim, Kyoo-Sung Noh, Joo-Yeoun Lee, "An Exploratory Study on Improvement Method of the Subway Congestion Based Big Data Convergence", JOURNAL OF DIGITAL CONVERGENCE, Vol.13, No.2, pp.35-42, 2015
- [7] <http://www.ict.r.or.kr>
- [8] Kyoung-Ho Choi, Jin-Ah Yoo, "A reviews on the social network analysis using R", Journal of the Korea Convergence Society, Vol. 6, No. 1, pp. 77-83, 2015.

- [9] James Manyika, Michael Chui, "Big data: The next frontier for innovation, competition, and productivity", McKinsey Global Institute, May 2011.
- [10] John Gantz, David Reinsel, "Extracting Value from Chaos", IDC IVIEW, June 2011.
- [11] Jeong-Mee Lee, "Understanding Big Data and Utilizing its Analysis into Library and Information Services", Journal of the Korean Biblia Society For Library And Information Science Vol. 24, No. 4, pp. 53-73, 2013.
- [12] Jung-Hoon Kim, Jun-Young Go, Keun-Ho Lee, "A Scheme of Social Engineering Attacks and Countermeasures Using Big Data based Conversion Voice Phishing", Journal of the Korea Convergence Society, Vol. 6, No. 1, pp. 85-91, 2015.
- [13] Keun-Won Kim, Dong-Woo Kim, Kyoo-Sung Noh, Joo-Yeoun Lee, "An Exploratory Study on Improvement Method of the Subway Congestion Based Big Data Convergence", Journal of Digital Convergence, Vol.13, No.2, pp.35-42, 2015.
- [14] Rencher, Alvin C., Christensen, William F., Methods of Multivariate Analysis, Wiley Series in Probability and Statistics, 709 (3rd ed.), John Wiley & Sons, 2012.
- [15] <http://kosis.kr/>
- [16] Hee-Seog Koh, A Study of the train operating efficiency - Incheon Subway Line 1 to attract utilizes the center, Korea National University of Transportation, 2015.

#### 김 진 수(Kim, Jin su)



- 1998년 2월 : 인천대학교 전자계산 공학과 (공학사)
- 2001년 8월 : 인하대학교 컴퓨터공학과 (공학석사)
- 2010년 2월 : 인하대학교 정보공학과 (공학박사)
- 2011년 3월 ~ 현재 : 안양대학교 교양학부대학 조교수

- 관심분야 : 데이터마이닝, 정보검색, 유비쿼터스 컴퓨팅, 무선 센서 네트워크, 빅데이터
- E-Mail : kjspace@anyang.ac.kr